

Beamforming Design for Full-Duplex Two-Way Amplify-and-Forward MIMO Relay

Yeonggyu Shim, *Student Member, IEEE*, Wan Choi, *Senior Member, IEEE*, and

Hyuncheol Park, *Senior Member, IEEE*

Abstract

We consider the full-duplex (FD) two-way amplify-and-forward relay system with imperfect cancellation of loopback self-interference (SI) and investigate joint design of relay and receive beamforming for minimizing the mean square error under a relay transmit power constraint. Due to loopback channel estimation error and limitation of analog-to-digital converter, the loopback SI cannot be completely canceled. Multiple antennas at the relay can help loopback SI suppression but beamforming is required to balance between the residual SI suppression and the desired signal transmission. Moreover, the relay beamforming matrix should be updated every time slot because the residual SI in the previous time slot is amplified by the current beamforming matrix and added to the received signals from the two sources in the current time slot. We derive the optimally balanced relay beamforming and receive beamforming matrices in closed form based on minimum mean square error, taking into account the propagation of the residual loopback SI from the first to the current time slot. We also propose beamforming design using only the channels of the m latest time slots, not from the first time slot. Based on our numerical results, we also identify when FD is beneficial and propose selection between FD and half-duplex according to signal-to-noise ratio and interference-to-noise ratio.

Y. Shim, W. Choi, and H. Park are with the School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Korea (e-mail: ygshim@kaist.ac.kr; wchoi@kaist.edu; hcpark@kaist.ac.kr).

I. INTRODUCTION

Two-way relaying systems achieve higher spectral efficiency than one-way relaying systems. In two-way relay channel (TWRC), two source nodes exchange messages with each other via a relay, using various multiple access broadcast (MABC) protocols consisting of the multiple access (MAC) phase and the broadcast (BC) phase [1]–[5]. Basically, in the MAC phase, two source nodes simultaneously send their signals to the relay node, and then the relay node amplifies and forwards the signal received in the MAC phase to the two source nodes in the BC phase. With a prior knowledge of its own transmitted signal, each source node is able to cancel the propagated self-interference (SI) back to it and decode the desired signal.

Because full-duplex (FD) systems allow concurrent transmission and reception in the same frequency band, they usually achieve higher spectral efficiency than half-duplex (HD) systems only if the loopback SI due to the concurrent transmission and reception is properly suppressed. Using a circulator which connects antenna to transceiver, a single antenna can be shared between the transmitter and the receiver in FD systems.

The merits of FD systems and two-way relaying systems in terms spectral efficiency motivated two-way relaying with a FD relay which outperforms both two-way relaying with a HD relay and one-way relaying with a FD relay. In [6]–[9], the two-way FD amplify-and-forward (AF) relay systems with two sources and one relay with a single antenna each were studied. Power allocation to maximize the achievable rate was explored in [6] but residual SI at each node was not considered assuming perfect SI suppression. However, given loopback channel estimation errors and limitation of analog-to-digital converter (ADC) in practical environments, the assumption of perfect SI suppression is too ideal to evaluate practical gains of two-way FD relaying. Contrary to [6], power allocation addressing residual SI due to imperfect SI cancelation was investigated

under a sum power constraint in [7]. To maximize the sum rate in the presence of residual SI, power allocation under individual power constraints and relay mode selection, among one-way HD, two-way HD, one-way FD, and two-way FD, were proposed in [8]. As another performance metric in the presence of residual SI, in [9], the outage probability of the two-way FD AF relay system was derived. The application of physical layer network coding to the two-way FD AF relay was investigated in terms of bit error rate [10] when there exists residual SI. A two-way FD AF relay system with multiple relays was addressed in [11] and an optimal relay selection scheme was proposed considering residual SI.

Although residual SI severely hurts the merit of FD, addressing the residual SI only with single antenna at a FD AF relay has limitations. On the other hand, beamforming with multiple antennas enables to suppress the residual SI, which motivated the study of a two-way FD AF relay system with a relay having multiple antennas. However, to the authors' best knowledge, there are only few studies on the two-way FD AF relay system with multiple antennas. In [12], a two-way FD AF relay system with multiple antennas was studied, where two sources have a single antenna each and the relay has multiple antennas. To maximize the achievable sum rate, beamforming at relay and power allocation at sources were jointly optimized with a zero forcing (ZF) constraint such that the residual loopback SI is perfectly canceled at the relay. However, a closed form solution was not available and thus an iterative algorithm was proposed in [12].

Contrary to the beamforming matrix which perfectly cancel out the residual SI with a ZF constraint, beamforming at a FD relay is required to balance between the residual SI suppression and the desired signal transmission. In this context, in a two-way FD AF relay system with multiple antennas with circulator at the relay as well as the two sources, we design the beamforming matrix balancing them in terms of minimum mean square error (MMSE) under a

power constraint at the relay. Because multiple antennas at the two sources are considered unlike [12], the beamforming matrix also need to be designed to address the inter-stream interference in the multi-input multi-output (MIMO) channels. Moreover, to optimally balance the residual SI suppression and the desired signal transmission, the beamforming matrix should be updated every time slot because the residual SI in the previous time slot is amplified by the current beamforming matrix and added to the received signals from the two sources in the current time slot.

In this paper, we analytically identify this coupled effect and show that the optimal beamforming matrix at each time slot is determined as a function of not only the current channels but also all the channels in the past time slots from the first transmission time slot. Using the Lagrangian method, we derive the optimized relay beamforming matrix and receive beamforming matrices at sources in closed form. To account for beamforming design with limited memory, we also propose a beamforming method using the channels in a limited number of the latest time slots, not from the first time slot, and show that the performance degradation is marginal relative to the beamforming method based on all the channel matrices from the first time slot to the current time slot. Finally, from our numerical results, we identify the signal-to-noise ratio (SNR) and interference-to-noise ratio (INR) region where the achievable sum rate of the proposed scheme is greater than that of the HD system. Based on this result, we propose a duplex mode selection scheme in terms of achievable sum rate according to SNR and INR.

The rest of this paper is organized as follows. Section II describes the system model. In Section III, we derive the optimal beamforming matrix in terms of MMSE and explore beamforming matrix design with limited memory size. In Section IV, we present numerical results. Finally, conclusions are drawn in Section V.

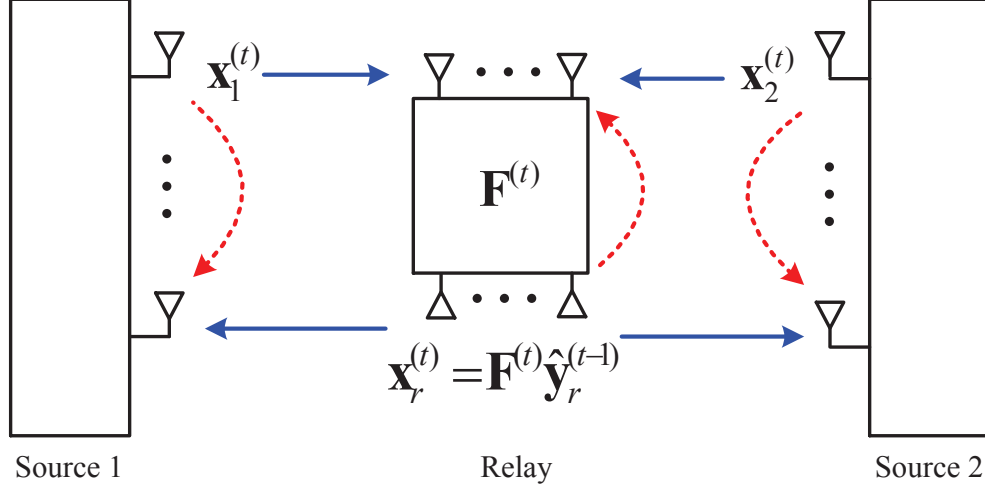


Fig. 1. Full-duplex two-way amplify-and-forward relay system in time slot t .

Notations: Matrices and vectors are denoted, respectively, by uppercase and lowercase boldface characters (e.g., \mathbf{A} and \mathbf{a}). The transpose, Hermitian, and inverse of \mathbf{A} are denoted, respectively, by \mathbf{A}^T , \mathbf{A}^H , and \mathbf{A}^{-1} . An $N \times N$ identity matrix and $N \times N$ matrix consisting of all zero entries are denoted, respectively, by \mathbf{I}_N and $\mathbf{0}_N$. The operators $\mathbb{E}[\cdot]$, $\text{tr}(\cdot)$, $\det(\cdot)$, $\text{vec}(\cdot)$, $\text{mat}(\cdot)$, \otimes , and \prod indicate the expectation, trace of a matrix, determinant of a matrix, matrix vectorization, inverse operation of $\text{vec}(\cdot)$, Kronecker product, and sequence product operators, respectively. Notations $\|\mathbf{a}\|$ and $\|\mathbf{A}\|_{\mathcal{F}}$ denotes 2-norm of \mathbf{a} and the Frobenius norm of \mathbf{A} , respectively.

II. SYSTEM MODEL

As shown in Fig. 1, consider a two-way FD AF relaying system consisting of two source nodes s_1 and s_2 with N_s antennas, and one relay node r with N_r antennas, where all nodes operate in the FD mode. In time slot 0, the two sources transmit their signals to the relay simultaneously. In time slot 1, the relay forwards its received signal after multiplying it by a beamforming matrix to the source nodes and two source nodes transmit their next signals to

the relay at the same time. In this manner, the two source nodes repeatedly exchange their information with each other via relay. In the t th time slot, source 1, source 2, and the relay transmit their signals $\mathbf{x}_1^{(t)} \in \mathbb{C}^{N_s \times 1}$, $\mathbf{x}_2^{(t)} \in \mathbb{C}^{N_s \times 1}$, and $\mathbf{x}_r^{(t)} \in \mathbb{C}^{N_r \times 1}$, respectively, of which covariance matrices are given by $\mathbb{E}[\mathbf{x}_i^{(t)} \mathbf{x}_i^{(t)H}] = p_i \mathbf{I}_{N_s}$ for $i \in \{1, 2\}$ and $\mathbb{E}[\mathbf{x}_r^{(t)} \mathbf{x}_r^{(t)H}] = p_r \mathbf{I}_{N_r}$. The noise signals at source 1, source 2, and the relay are denoted as $\mathbf{n}_i^{(t)} \in \mathbb{C}^{N_s \times 1}$ for $i \in \{1, 2\}$ and $\mathbf{n}_r^{(t)} \in \mathbb{C}^{N_r \times 1}$, respectively, and are assumed to be zero-mean white Gaussian noise (AWGN) with $\mathbb{E}[\mathbf{n}_i^{(t)} \mathbf{n}_i^{(t)H}] = \sigma_{n,i}^2 \mathbf{I}_{N_s}$ for $i \in \{1, 2\}$ and $\mathbb{E}[\mathbf{n}_r^{(t)} \mathbf{n}_r^{(t)H}] = \sigma_{n,r}^2 \mathbf{I}_{N_r}$.

The channel links are modeled as independent and frequency-flat fading channels and assumed to be static in each time slot. In time slot t , the channel matrices between source 1 and the relay, source 2 and the relay, the relay and source 1, and the relay and source 2 are represented as $\mathbf{H}_{1,r}^{(t)}, \mathbf{H}_{2,r}^{(t)} \in \mathbb{C}^{N_r \times N_s}$ and $\mathbf{H}_{r,1}^{(t)}, \mathbf{H}_{r,2}^{(t)} \in \mathbb{C}^{N_s \times N_r}$, respectively. Also, the loopback channel matrices at source 1, source 2, and the relay are represented as $\mathbf{H}_{i,i}^{(t)} \in \mathbb{C}^{N_s \times N_s}$ for $i \in \{1, 2\}$ and $\mathbf{H}_{r,r}^{(t)} \in \mathbb{C}^{N_r \times N_r}$, respectively.

The channel state information (CSI) of the loopback channels is assumed to be imperfect due to channel estimation errors, whereas, to focus only on the residual loopback SI related to full-duplex operation, CSI between two nodes is assumed to be perfectly known at each node. Since each node knows its own transmitted signals and perfect CSI between two nodes, the backward propagated SI can be canceled perfectly. However, due to imperfect CSI of the loopback channels, each node cannot perfectly cancel their loopback SI.

In a FD relay system, a loopback channel exists at each node. The imperfect loopback channel estimation results in loopback SI, but the imperfect backward channel (i.e., channel between two sources via relay) estimation errors are not related to loopback SI. To focus on the effect of loopback SI in FD relay systems, we consider imperfect loopback channel estimation but perfect

backward channel estimation is assumed. Moreover, as the number of involved time slots increase, the impact of the loopback channel estimation errors dominates that of the backward channel estimation errors.

The relationship between the exact channel and the estimated channel is modeled as $\mathbf{H}_{i,i}^{(t)} = \hat{\mathbf{H}}_{i,i}^{(t)} + \Delta_{i,i}^{(t)}$ for $i \in \{1, 2, r\}$, where $\hat{\mathbf{H}}_{j,j}^{(t)} \in \mathbb{C}^{N_s \times N_s}$ for $j \in \{1, 2\}$ and $\hat{\mathbf{H}}_{r,r}^{(t)} \in \mathbb{C}^{N_r \times N_r}$ are the estimated channel matrices, and $\Delta_{j,j}^{(t)} \in \mathbb{C}^{N_s \times N_s}$ and $\Delta_{r,r}^{(t)} \in \mathbb{C}^{N_r \times N_r}$ are channel estimation error matrices. It is assumed that $\Delta_{i,i}^{(t)}$ is Gaussian distributed with zero mean, and $\mathbb{E}[\text{vec}(\Delta_{j,j}^{(t)})\text{vec}(\Delta_{j,j}^{(t)H})] = \sigma_{e,j}^2 \mathbf{I}_{N_j^2}$ and $\mathbb{E}[\text{vec}(\Delta_{r,r}^{(t)})\text{vec}(\Delta_{r,r}^{(t)H})] = \sigma_{e,r}^2 \mathbf{I}_{N_r^2}$. Also, $\Delta_{i,i}^{(t)}$ for $t \geq 1$ are independent of each other.

The transmission process starts from time slot 0, so the sources transmit their signals but the relay do not in time slot 0. Because the relay starts transmitting from time slot 1, FD operation substantially starts from time slot 1. In time slot 0, the two source nodes simultaneously broadcast their signals $\mathbf{x}_1^{(0)} \in \mathbb{C}^{N_s \times 1}$ and $\mathbf{x}_2^{(0)} \in \mathbb{C}^{N_s \times 1}$ to the relay and then the received signal $\mathbf{y}_r^{(0)} \in \mathbb{C}^{N_r \times 1}$ at the relay becomes

$$\mathbf{y}_r^{(0)} = \mathbf{H}_{1,r}^{(0)}\mathbf{x}_1^{(0)} + \mathbf{H}_{2,r}^{(0)}\mathbf{x}_2^{(0)} + \mathbf{n}_r^{(0)}. \quad (1)$$

In time slot 1, the relay forwards its received signal to the source nodes after multiplying it by a relay beamforming matrix $\mathbf{F}^{(1)} \in \mathbb{C}^{N_r \times N_r}$ and then the transmit signal $\mathbf{x}_r^{(1)} \in \mathbb{C}^{N_r \times 1}$ at the relay is represented as

$$\mathbf{x}_r^{(1)} = \mathbf{F}^{(1)}\mathbf{y}_r^{(0)} \quad (2)$$

of which transmit power is given by $\mathbb{E}[||\mathbf{x}_r^{(1)}||^2] = N_r p_r$. With imperfect loopback SI cancellation, the received signal at the relay in time slot 1 is given by

$$\hat{\mathbf{y}}_r^{(1)} = \mathbf{H}_{1,r}^{(1)}\mathbf{x}_1^{(1)} + \mathbf{H}_{2,r}^{(1)}\mathbf{x}_2^{(1)} + \Delta_{r,r}^{(1)}\mathbf{x}_r^{(1)} + \mathbf{n}_r^{(1)}. \quad (3)$$

Substituting (2) into (3) leads to

$$\hat{\mathbf{y}}_r^{(1)} = \mathbf{H}_{1,r}^{(1)} \mathbf{x}_1^{(1)} + \mathbf{H}_{2,r}^{(1)} \mathbf{x}_2^{(1)} + \mathbf{n}_r^{(1)} + \Delta_{r,r}^{(1)} \mathbf{F}^{(1)} (\mathbf{H}_{1,r}^{(0)} \mathbf{x}_1^{(0)} + \mathbf{H}_{2,r}^{(0)} \mathbf{x}_2^{(0)} + \mathbf{n}_r^{(0)}). \quad (4)$$

At source l in time slot 1, subtracting the propagated SI back to the source node $\mathbf{H}_{r,l}^{(1)} \mathbf{F}^{(1)} \mathbf{H}_{l,r}^{(0)} \mathbf{x}_l^{(0)}$, and the loopback SI $\hat{\mathbf{H}}_{l,l}^{(1)} \mathbf{x}_l^{(1)}$, the received signal at source l is given by

$$\hat{\mathbf{y}}_l^{(1)} = \mathbf{H}_{r,l}^{(1)} \mathbf{F}^{(1)} \mathbf{H}_{\bar{l},r}^{(0)} \mathbf{x}_{\bar{l}}^{(0)} + \mathbf{H}_{r,l}^{(1)} \mathbf{F}^{(1)} \mathbf{n}_r^{(0)} + \Delta_{l,l}^{(1)} \mathbf{x}_l^{(1)} + \mathbf{n}_l^{(1)}, \quad (5)$$

where \bar{l} means the index of the other source (i.e., $\bar{l} = 2$ if $l = 1$ and $\bar{l} = 1$ if $l = 2$). Source l multiplies $\hat{\mathbf{y}}_l^{(1)}$ by a receive beamforming matrix $\mathbf{R}_l^{(1)} \in \mathbb{C}^{N_s \times N_s}$.

Accordingly, considering imperfect loopback SI cancellation, the received signal at the relay in time slot $t - 1$ ($t \geq 2$) is given by

$$\hat{\mathbf{y}}_r^{(t-1)} = \mathbf{H}_{1,r}^{(t-1)} \mathbf{x}_1^{(t-1)} + \mathbf{H}_{2,r}^{(t-1)} \mathbf{x}_2^{(t-1)} + \mathbf{n}_r^{(t-1)} + \sum_{i=0}^{t-2} \left\{ \prod_{j=1}^{t-1-i} (\Delta_{r,r}^{(t-j)} \mathbf{F}^{(t-j)}) (\mathbf{H}_{1,r}^{(i)} \mathbf{x}_1^{(i)} + \mathbf{H}_{2,r}^{(i)} \mathbf{x}_2^{(i)} + \mathbf{n}_r^{(i)}) \right\} \quad (6)$$

which shows that the current received signal is affected by the beamforming matrices in the past time slots. In time slot t , the forwarded signal from the relay to the source nodes is given by

$$\mathbf{x}_r^{(t)} = \mathbf{F}^{(t)} \hat{\mathbf{y}}_r^{(t-1)} \quad (7)$$

where $\mathbf{F}^{(t)} \in \mathbb{C}^{N_r \times N_r}$ is the relay beamforming matrix at time slot t and $\mathbf{x}_r^{(t)} \in \mathbb{C}^{N_r \times 1}$. Subtracting the propagated SI back to the source node $\mathbf{H}_{r,l}^{(t)} \mathbf{F}^{(t)} \mathbf{H}_{l,r}^{(t-1)} \mathbf{x}_l^{(t-1)}$ and the loopback SI $\hat{\mathbf{H}}_{l,l}^{(t)} \mathbf{x}_l^{(t)}$, the received signal at source l in time slot t ($t \geq 2$) is obtained as

$$\begin{aligned} \hat{\mathbf{y}}_l^{(t)} &= \mathbf{H}_{r,l}^{(t)} \mathbf{F}^{(t)} \mathbf{H}_{\bar{l},r}^{(t-1)} \mathbf{x}_{\bar{l}}^{(t-1)} + \mathbf{H}_{r,l}^{(t)} \mathbf{F}^{(t)} \sum_{i=0}^{t-2} \left\{ \prod_{j=1}^{t-1-i} (\Delta_{r,r}^{(t-j)} \mathbf{F}^{(t-j)}) (\mathbf{H}_{1,r}^{(i)} \mathbf{x}_1^{(i)} + \mathbf{H}_{2,r}^{(i)} \mathbf{x}_2^{(i)} + \mathbf{n}_r^{(i)}) \right\} \\ &\quad + \mathbf{H}_{r,l}^{(t)} \mathbf{F}^{(t)} \mathbf{n}_r^{(t-1)} + \Delta_{l,l}^{(t)} \mathbf{x}_l^{(t)} + \mathbf{n}_l^{(t)}. \end{aligned} \quad (8)$$

After subtracting SI, source l multiplies $\hat{\mathbf{y}}_l^{(t)}$ by a receive beamforming matrix $\mathbf{R}_l^{(t)} \in \mathbb{C}^{N_s \times N_s}$.

III. MMSE BASED BEAMFORMING DESIGN

The problem for joint design of relay beamforming matrices and source receive beamforming matrices is non-convex and basically intractable. Therefore, we propose an iterative algorithm which decouples the primal problem into two subproblems and solve them alternately; one is for relay beamforming design and the other is for receive beamforming design at sources.

As identified in (6), the current received signal is a function of the beamforming matrices, the channel estimation error matrices, and transmitted signals at source in the past time slots, which implies that to optimally balance the residual SI suppression and the desired signal transmission, the beamforming matrix should be updated every time slot. Without loss of generality, we design the relay beamforming matrix based on the m latest time slots; if $m = t - 1$, the beamforming matrix in time slot t is based on all the time slots from the first time slot. Otherwise if $m < t - 1$, the beamforming design corresponds to the scenario when the memory size at the relay is limited. Unlike one-way relay systems, transmitted signals at each source during the past time slots become interference signals at each source in the current time slot in two-way relay systems. In addition, the relay transmits the received noise signals during the past time slots to the sources. Therefore, the minimum mean square error (MMSE) criterion is an effective metric enabling a compromise between interference suppression and noise reduction. Moreover, the MMSE criterion allows tractable analysis and optimization of the two-way FD AF relay system. For similar reasons, mean square error (MSE) has been popularly used for beamforming design in two-way relay systems in the literature such as [13]–[18].

To balance the residual SI suppression and the desired signal transmission, the relay beamforming matrix $\mathbf{F}^{(t)}$ and receive beamforming matrices $\mathbf{R}_i^{(t)}$ for $i \in \{1, 2\}$ in time slot t (≥ 1) are determined based on the MSE criterion under a transmit power constraint at the relay. The

corresponding optimization problem is formulated as

$$\mathbb{P}^{(t)} : \min_{\mathbf{F}^{(t)}, \alpha^{(t)}, \mathbf{R}_1^{(t)}, \mathbf{R}_2^{(t)}} J(\mathbf{F}^{(t)}, \alpha^{(t)}, \mathbf{R}_1^{(t)}, \mathbf{R}_2^{(t)}) \quad \text{s.t.} \quad \mathbb{E}[\|\mathbf{x}_r^{(t)}\|^2] = N_r p_r \quad (9)$$

where the sum of MSE $J(\mathbf{F}^{(t)}, \alpha^{(t)}, \mathbf{R}_1^{(t)}, \mathbf{R}_2^{(t)})$ is defined as

$$J(\mathbf{F}^{(t)}, \alpha^{(t)}, \mathbf{R}_1^{(t)}, \mathbf{R}_2^{(t)}) \triangleq \mathbb{E}[\|\mathbf{x}_2^{(t-1)} - \alpha^{(t-1)} \mathbf{R}_1^{(t)H} \hat{\mathbf{y}}_1^{(t)}\|^2] + \mathbb{E}[\|\mathbf{x}_1^{(t-1)} - \alpha^{(t-1)} \mathbf{R}_2^{(t)H} \hat{\mathbf{y}}_2^{(t)}\|^2] \quad (10)$$

and $\alpha^{(t-1)}$ is a positive scaling factor. Note that $\hat{\mathbf{y}}_l^{(t)}$ is multiplied by the positive scaling factor $\alpha^{(t-1)}$ since the relay beamforming matrix $\mathbf{F}^{(t)}$ is split into two components, as in [15], [19], such that $\mathbf{F}^{(t)} = \alpha^{(t)} \bar{\mathbf{F}}^{(t)}$ to facilitate the derivation of a closed form solution, where $\|\bar{\mathbf{F}}^{(t)}\|_{\mathcal{F}} = 1$ and thus $\alpha^{(t)}$ and $\bar{\mathbf{F}}^{(t)}$ represent power amplification and steering directions, respectively.

The Lagrangian of the optimization problem is defined as

$$\begin{aligned} \mathcal{L}(\mathbf{F}^{(t)}, \alpha^{(t)}, \mathbf{R}_1^{(t)}, \mathbf{R}_2^{(t)}, \lambda^{(t)}) \\ &= J(\mathbf{F}^{(t)}, \alpha^{(t)}, \mathbf{R}_1^{(t)}, \mathbf{R}_2^{(t)}) + \lambda^{(t)} (\mathbb{E}[\|\mathbf{x}_r^{(t)}\|^2] - N_r p_r) \\ &= N_s(p_1 + p_2) - \alpha^{(t-1)} \{\text{tr}(\mathbf{W}_{f,0}^{(t)H} \mathbf{F}^{(t)}) + \text{tr}(\mathbf{W}_{f,0}^{(t)} \mathbf{F}^{(t)H})\} \\ &\quad + \alpha^{(t-2)} \{\mathbb{E}[\|\mathbf{R}_1^{(t)H} \hat{\mathbf{y}}_1^{(t)}\|^2] + \mathbb{E}[\|\mathbf{R}_2^{(t)H} \hat{\mathbf{y}}_2^{(t)}\|^2]\} + \lambda^{(t)} (\mathbb{E}[\|\mathbf{x}_r^{(t)}\|^2] - N_r p_r), \end{aligned} \quad (11)$$

where $\lambda^{(t)}$ is a Lagrangian multiplier in time slot t , $\mathbf{W}_{f,0}^{(t)} = p_1 \mathbf{H}_{r,2}^{(t)H} \mathbf{R}_2^{(t)} \mathbf{H}_{1,r}^{(t-1)H} + p_2 \mathbf{H}_{r,1}^{(t)H} \mathbf{R}_1^{(t)} \mathbf{H}_{2,r}^{(t-1)H}$,

$$\begin{aligned} \mathbb{E}[\|\mathbf{R}_1^{(t)H} \hat{\mathbf{y}}_1^{(t)}\|^2] &= \mathbb{E}[\text{tr}\{\mathbf{R}_1^{(t)H} \mathbf{H}_{r,1}^{(t)} \mathbf{F}^{(t)} (\mathbf{G}_0^{(t)} + p_2 \mathbf{H}_{2,r}^{(t-1)} \mathbf{H}_{2,r}^{(t-1)H} + \sigma_{n,r}^2 \mathbf{I}_{N_r}) \mathbf{F}^{(t)H} \mathbf{H}_{r,1}^{(t)H} \mathbf{R}_1^{(t)}\}] \\ &\quad + p_1 \mathbb{E}[\text{tr}(\mathbf{R}_1^{(t)H} \Delta_{1,1}^{(t)} \Delta_{1,1}^{(t)H} \mathbf{R}_1^{(t)})] + \sigma_{n,1}^2 \text{tr}(\mathbf{R}_1^{(t)H} \mathbf{R}_1^{(t)}), \\ \mathbb{E}[\|\mathbf{R}_2^{(t)H} \hat{\mathbf{y}}_2^{(t)}\|^2] &= \mathbb{E}[\text{tr}\{\mathbf{R}_2^{(t)H} \mathbf{H}_{r,2}^{(t)} \mathbf{F}^{(t)} (\mathbf{G}_0^{(t)} + p_1 \mathbf{H}_{1,r}^{(t-1)} \mathbf{H}_{1,r}^{(t-1)H} + \sigma_{n,r}^2 \mathbf{I}_{N_r}) \mathbf{F}^{(t)H} \mathbf{H}_{r,2}^{(t)H} \mathbf{R}_2^{(t)}\}] \\ &\quad + p_2 \mathbb{E}[\text{tr}(\mathbf{R}_2^{(t)H} \Delta_{2,2}^{(t)} \Delta_{2,2}^{(t)H} \mathbf{R}_2^{(t)})] + \sigma_{n,2}^2 \text{tr}(\mathbf{R}_2^{(t)H} \mathbf{R}_2^{(t)}), \\ \mathbb{E}[\|\mathbf{x}_r^{(t)}\|^2] &= \mathbb{E}[\text{tr}\{\mathbf{F}^{(t)} (\mathbf{G}_0^{(t)} + p_1 \mathbf{H}_{1,r}^{(t-1)} \mathbf{H}_{1,r}^{(t-1)H} + p_2 \mathbf{H}_{2,r}^{(t-1)} \mathbf{H}_{2,r}^{(t-1)H} + \sigma_{n,r}^2 \mathbf{I}_{N_r}) \mathbf{F}^{(t)H}\}], \end{aligned} \quad (12)$$

where $\mathbf{G}_0^{(1)} = \mathbf{0}_{N_r}$,

$$\mathbf{G}_0^{(t)} = \sum_{i=0}^{t-2} \left\{ \prod_{j=1}^{t-1-i} (\Delta_{r,r}^{(t-j)} \mathbf{F}^{(t-j)}) (p_1 \mathbf{H}_{1,r}^{(i)} \mathbf{H}_{1,r}^{(i)H} + p_2 \mathbf{H}_{2,r}^{(i)} \mathbf{H}_{2,r}^{(i)H} + \sigma_{n,r}^2 \mathbf{I}_{N_r}) \prod_{j=1}^{t-1-i} (\mathbf{F}^{(i+j)H} \Delta_{r,r}^{(i+j)H}) \right\}$$

for $2 \leq t \leq m+1$, and

$$\begin{aligned} \mathbf{G}_0^{(t)} = & \sum_{i=0}^{t-2-m} \left\{ \prod_{j=1}^m (\Delta_{r,r}^{(t-j)} \mathbf{F}^{(t-j)}) \prod_{j=m+1}^{t-1-i} (\Delta_{r,r}^{(t-j)} \mathbf{F}^{(t-m)}) \right. \\ & (p_1 \mathbf{H}_{1,r}^{(t-1-m)} \mathbf{H}_{1,r}^{(t-1-m)H} + p_2 \mathbf{H}_{2,r}^{(t-1-m)} \mathbf{H}_{2,r}^{(t-1-m)H} + \sigma_{n,r}^2 \mathbf{I}_{N_r}) \\ & \left. \prod_{j=1}^{t-1-m-i} (\mathbf{F}^{(t-m)H} \Delta_{r,r}^{(i+j)H}) \prod_{j=t-m}^{t-1} (\mathbf{F}^{(j)H} \Delta_{r,r}^{(j)H}) \right\} \\ & + \sum_{i=t-1-m}^{t-2} \left\{ \prod_{j=1}^{t-1-i} (\Delta_{r,r}^{(t-j)} \mathbf{F}^{(t-j)}) (p_1 \mathbf{H}_{1,r}^{(i)} \mathbf{H}_{1,r}^{(i)H} + p_2 \mathbf{H}_{2,r}^{(i)} \mathbf{H}_{2,r}^{(i)H} + \sigma_{n,r}^2 \mathbf{I}_{N_r}) \prod_{j=1}^{t-1-i} (\mathbf{F}^{(i+j)H} \Delta_{r,r}^{(i+j)H}) \right\} \end{aligned}$$

for $t \geq m+2$.

(13)

Note that m is the number of past time slots considered in the beamforming matrix design and the expectations in (12) are with respect to $\Delta_{i,i}^{(t)}$ for $i \in \{1, 2, r\}$.

Lemma 1: Let Δ_j be an $N \times N$ random matrix with $\mathbb{E}[\text{vec}(\Delta_j) \text{vec}(\Delta_j)^H] = \sigma^2 \mathbf{I}_{N^2}$ and Δ_j for $j \geq 1$ are independent of each other.

$$\mathbb{E} \left[\text{tr} \left\{ \mathbf{V}_v \prod_{j=1}^{v-1} (\Delta_{v-j} \mathbf{V}_{v-j}) \prod_{j=1}^{v-1} (\mathbf{V}_j^H \Delta_j^H \mathbf{V}_j^H) \right\} \right] = (\sigma^2)^{v-1} \prod_{j=1}^v \text{tr}(\mathbf{V}_j \mathbf{V}_j^H) \text{ for } v \geq 2. \quad (14)$$

Proof: We prove this by induction. From *Lemma 1* in [20], (14) holds for $v = 2$. That is, $\mathbb{E}[\text{tr}\{\mathbf{V}_2 \Delta_1 \mathbf{V}_1 \mathbf{V}_1^H \Delta_1^H \mathbf{V}_2^H\}] = \sigma^2 \text{tr}(\mathbf{V}_1 \mathbf{V}_1^H) \text{tr}(\mathbf{V}_2 \mathbf{V}_2^H)$. Assume (14) holds when $v = k-1$ and then we prove that it also holds when $v = k$.

When $v = k$, the left hand side of (14) is rewritten as

LHS of (14)

$$\stackrel{(a)}{=} \mathbb{E} \left[\text{tr} \left\{ \prod_{j=1}^{k-1} (\Delta_{k-j} \mathbf{V}_{k-j}) \prod_{j=1}^{k-1} (\mathbf{V}_j^H \Delta_j^H) \mathbf{V}_k^H \mathbf{V}_k \right\} \right] \quad (15)$$

$$\stackrel{(b)}{=} \mathbb{E} \left[\text{vec}(\Delta_{k-1})^H \left\{ \left(\mathbf{V}_{k-1} \prod_{j=1}^{k-2} (\Delta_{k-1-j} \mathbf{V}_{k-1-j}) \prod_{j=1}^{k-2} (\mathbf{V}_j^H \Delta_j^H) \mathbf{V}_{k-1}^H \right)^T \otimes (\mathbf{V}_k^H \mathbf{V}_k) \right\} \text{vec}(\Delta_{k-1}) \right] \quad (16)$$

$$\stackrel{(c)}{=} \text{tr} \left(\mathbb{E} \left[\text{vec}(\Delta_{k-1})^H \left\{ \left(\mathbf{V}_{k-1} \prod_{j=1}^{k-2} (\Delta_{k-1-j} \mathbf{V}_{k-1-j}) \prod_{j=1}^{k-2} (\mathbf{V}_j^H \Delta_j^H) \mathbf{V}_{k-1}^H \right)^T \otimes (\mathbf{V}_k^H \mathbf{V}_k) \right\} \text{vec}(\Delta_{k-1}) \right] \right) \quad (17)$$

$$\stackrel{(d)}{=} \text{tr} \left\{ \mathbb{E} \left[\left(\mathbf{V}_{k-1} \prod_{j=1}^{k-2} (\Delta_{k-1-j} \mathbf{V}_{k-1-j}) \prod_{j=1}^{k-2} (\mathbf{V}_j^H \Delta_j^H) \mathbf{V}_{k-1}^H \right)^T \otimes (\mathbf{V}_k^H \mathbf{V}_k) \right] \mathbb{E} \left[\text{vec}(\Delta_{k-1}) \text{vec}(\Delta_{k-1})^H \right] \right\} \quad (18)$$

$$\stackrel{(e)}{=} \sigma^2 \text{tr}(\mathbf{V}_k \mathbf{V}_k^H) \mathbb{E} \left[\text{tr} \left\{ \mathbf{V}_{k-1} \prod_{j=1}^{k-2} (\Delta_{k-1-j} \mathbf{V}_{k-1-j}) \prod_{j=1}^{k-2} (\mathbf{V}_j^H \Delta_j^H) \mathbf{V}_{k-1}^H \right\} \right] \quad (19)$$

where (a) is due to $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$; (b) is due to $\text{tr}(\mathbf{ABCD}) = \text{vec}(\mathbf{C}^T)^T (\mathbf{B}^T \otimes \mathbf{D}) \text{vec}(\mathbf{A})$; (c) is because $c = \text{tr}(c)$ when c is a constant; (d) follows from $\text{tr}(\mathbb{E}[\mathbf{A}]) = \mathbb{E}[\text{tr}(\mathbf{A})]$, $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$, and independence of Δ_j ; (e) holds because $\text{tr}(\mathbb{E}[\mathbf{A}]) = \mathbb{E}[\text{tr}(\mathbf{A})]$, $\mathbb{E}[\text{vec}(\Delta_{k-1}) \text{vec}(\Delta_{k-1})^H] = \sigma^2 \mathbf{I}_{N^2}$, $\text{tr}(\mathbf{A} \otimes \mathbf{B}) = \text{tr}(\mathbf{A}) \text{tr}(\mathbf{B})$, and $\text{tr}(\mathbf{A})^T = \text{tr}(\mathbf{A})$.

Finally, using the induction hypothesis that (14) holds for $k - 1$, (19) becomes

$$(\sigma^2)^{k-1} \prod_{j=1}^k \text{tr}(\mathbf{V}_j \mathbf{V}_j^H) \quad (20)$$

■

Remark: Lemma 1 provides a mathematical tool to reach the main result. Specifically, to calculate MSE, the expectations on the received signals have to be calculated and Lemma 1 enables to calculate the expectations since the received signals are constituted by the product

of unknown random matrices (i.e., channel estimation error matrices) and known deterministic matrices (i.e., channel matrices and relay beamforming matrices) in rotation, which is the form in *Lemma 1*. Consequently, *Lemma 1* can be applied to the system design in which the received signal is presented by the product of random matrices and deterministic matrices in rotation.

Using *Lemma 1*, $\mathbb{E}[\|\mathbf{R}_1^{(t)} \hat{\mathbf{y}}_1^{(t)}\|^2]$, $\mathbb{E}[\|\mathbf{R}_2^{(t)} \hat{\mathbf{y}}_2^{(t)}\|^2]$, and $\mathbb{E}[\|\mathbf{x}_r^{(t)}\|^2]$ are obtained, respectively, as

$$\begin{aligned} \mathbb{E}[\|\mathbf{R}_1^{(t)} \hat{\mathbf{y}}_1^{(t)}\|^2] &= \text{tr}\{\mathbf{H}_{r,1}^{(t)H} \mathbf{R}_1^{(t)} \mathbf{R}_1^{(t)H} \mathbf{H}_{r,1}^{(t)} \mathbf{F}^{(t)} (\mathbf{G}_c^{(t)} + p_2 \mathbf{H}_{2,r}^{(t-1)} \mathbf{H}_{2,r}^{(t-1)H} + \sigma_{n,r}^2 \mathbf{I}_{N_r}) \mathbf{F}^{(t)H}\} \\ &\quad + (N_s p_1 \sigma_{e,1}^2 + \sigma_{n,1}^2) \text{tr}(\mathbf{R}_1^{(t)H} \mathbf{R}_1^{(t)}), \\ \mathbb{E}[\|\mathbf{R}_2^{(t)} \hat{\mathbf{y}}_2^{(t)}\|^2] &= \text{tr}\{\mathbf{H}_{r,2}^{(t)H} \mathbf{R}_2^{(t)} \mathbf{R}_2^{(t)H} \mathbf{H}_{r,2}^{(t)} \mathbf{F}^{(t)} (\mathbf{G}_c^{(t)} + p_1 \mathbf{H}_{1,r}^{(t-1)} \mathbf{H}_{1,r}^{(t-1)H} + \sigma_{n,r}^2 \mathbf{I}_{N_r}) \mathbf{F}^{(t)H}\} \\ &\quad + (N_s p_2 \sigma_{e,2}^2 + \sigma_{n,2}^2) \text{tr}(\mathbf{R}_2^{(t)H} \mathbf{R}_2^{(t)}), \\ \mathbb{E}[\|\mathbf{x}_r^{(t)}\|^2] &= \text{tr}\{\mathbf{F}^{(t)} (\mathbf{G}_c^{(t)} + p_1 \mathbf{H}_{1,r}^{(t-1)} \mathbf{H}_{1,r}^{(t-1)H} + p_2 \mathbf{H}_{2,r}^{(t-1)} \mathbf{H}_{2,r}^{(t-1)H} + \sigma_{n,r}^2 \mathbf{I}_{N_r}) \mathbf{F}^{(t)H}\}, \end{aligned} \quad (21)$$

where

$$\begin{aligned} &\mathbf{G}_c^{(t)} \\ &= \Gamma_1 \sigma_{e,r}^2 \mathbf{I}_{N_r} \text{tr}\{\mathbf{F}^{(t-1)} (p_1 \mathbf{H}_{1,r}^{(t-2)} \mathbf{H}_{1,r}^{(t-2)H} + p_2 \mathbf{H}_{2,r}^{(t-2)} \mathbf{H}_{2,r}^{(t-2)H} + \sigma_{n,r}^2 \mathbf{I}_{N_r}) \mathbf{F}^{(t-1)H}\} \\ &+ \Gamma_2 \sum_{i=2}^{\min(m, t-1)} \left[(\sigma_{e,r}^2)^i \mathbf{I}_{N_r} \prod_{j=t+1-i}^{t-1} \text{tr}(\mathbf{F}^{(j)} \mathbf{F}^{(j)H}) \right. \\ &\quad \left. \text{tr}\{\mathbf{F}^{(t-i)} (p_1 \mathbf{H}_{1,r}^{(t-i)} \mathbf{H}_{1,r}^{(t-i)H} + p_2 \mathbf{H}_{2,r}^{(t-i)} \mathbf{H}_{2,r}^{(t-i)H} + \sigma_{n,r}^2 \mathbf{I}_{N_r}) \mathbf{F}^{(t-i)H}\} \right] \\ &+ \Gamma_3 \sum_{i=2}^{t-1} \left[(\sigma_{e,r}^2)^i \mathbf{I}_{N_r} \prod_{j=t+1-i}^{t-1} \text{tr}(\mathbf{F}^{(j)} \mathbf{F}^{(j)H}) \right. \\ &\quad \left. \text{tr}\{\mathbf{F}^{(t-i)} (p_1 \mathbf{H}_{1,r}^{(t-1-i)} \mathbf{H}_{1,r}^{(t-1-i)H} + p_2 \mathbf{H}_{2,r}^{(t-1-i)} \mathbf{H}_{2,r}^{(t-1-i)H} + \sigma_{n,r}^2 \mathbf{I}_{N_r}) \mathbf{F}^{(t-i)H}\} \right] \end{aligned} \quad (22)$$

and Γ_1 , Γ_2 , and Γ_3 are defined as

$$\begin{aligned}
\Gamma_1 &= \Gamma_2 = \Gamma_3 = \mathbf{0}_{N_r} \text{ for } t = 1 \\
\Gamma_1 &= \mathbf{I}_{N_r} \text{ and } \Gamma_2 = \Gamma_3 = \mathbf{0}_{N_r} \text{ for } t = 2 \\
\Gamma_1 &= \Gamma_3 = \mathbf{I}_{N_r} \text{ and } \Gamma_2 = \mathbf{0}_{N_r} \text{ for } t \geq 3 \text{ and } m = 1 \\
\Gamma_1 &= \Gamma_2 = \mathbf{I}_{N_r} \text{ and } \Gamma_3 = \mathbf{0}_{N_r} \text{ for } 3 \leq t \leq m + 1 \\
\Gamma_1 &= \Gamma_2 = \Gamma_3 = \mathbf{I}_{N_r} \text{ for } t \geq m + 2
\end{aligned} \tag{23}$$

A. Relay beamforming design

The problem of relay beamforming matrix design in each time slot is formulated as

$$\mathbb{P}_f^{(t)} : \min_{\mathbf{F}^{(t)}, \alpha^{(t)}} J(\mathbf{F}^{(t)}, \alpha^{(t)}) \quad \text{s.t.} \quad \mathbb{E}[\|\mathbf{x}_r^{(t)}\|^2] = N_r p_r \tag{24}$$

Using (21) and substituting $\mathbf{F}^{(t)} = \alpha^{(t)} \bar{\mathbf{F}}^{(t)}$, the Lagrangian in (11) is reduced to

$$\begin{aligned}
\mathcal{L}(\bar{\mathbf{F}}^{(t)}, \alpha^{(t)}, \lambda^{(t)}) &= N_s(p_1 + p_2) - \{\text{tr}(\mathbf{W}_{f,0}^{(t)H} \bar{\mathbf{F}}^{(t)}) + \text{tr}(\mathbf{W}_{f,0}^{(t)} \bar{\mathbf{F}}^{(t)H})\} \\
&\quad + \text{tr}(\mathbf{W}_{f,1}^{(t)} \bar{\mathbf{F}}^{(t)} \mathbf{G}_1^{(t)} \bar{\mathbf{F}}^{(t)H}) + \text{tr}(\mathbf{W}_{f,2}^{(t)} \bar{\mathbf{F}}^{(t)} \mathbf{G}_2^{(t)} \bar{\mathbf{F}}^{(t)H}) \\
&\quad + \alpha^{(t)-2} w_f^{(t)} + \lambda^{(t)} \{\alpha^{(t)2} \text{tr}(\bar{\mathbf{F}}^{(t)} \mathbf{G}_r^{(t)} \bar{\mathbf{F}}^{(t)H}) - N_r p_r\},
\end{aligned} \tag{25}$$

where

$$\begin{aligned}
\mathbf{G}_1^{(t)} &= \mathbf{G}_c^{(t)} + p_2 \mathbf{H}_{2,r}^{(t-1)} \mathbf{H}_{2,r}^{(t-1)H} + \sigma_{n,r}^2 \mathbf{I}_{N_r}, \\
\mathbf{G}_2^{(t)} &= \mathbf{G}_c^{(t)} + p_1 \mathbf{H}_{1,r}^{(t-1)} \mathbf{H}_{1,r}^{(t-1)H} + \sigma_{n,r}^2 \mathbf{I}_{N_r}, \\
\mathbf{G}_r^{(t)} &= \mathbf{G}_c^{(t)} + p_1 \mathbf{H}_{1,r}^{(t-1)} \mathbf{H}_{1,r}^{(t-1)H} + p_2 \mathbf{H}_{2,r}^{(t-1)} \mathbf{H}_{2,r}^{(t-1)H} + \sigma_{n,r}^2 \mathbf{I}_{N_r}, \\
\mathbf{W}_{f,1}^{(t)} &= \mathbf{H}_{r,1}^{(t)H} \mathbf{R}_1^{(t)} \mathbf{R}_1^{(t)H} \mathbf{H}_{r,1}^{(t)}, \\
\mathbf{W}_{f,2}^{(t)} &= \mathbf{H}_{r,2}^{(t)H} \mathbf{R}_2^{(t)} \mathbf{R}_2^{(t)H} \mathbf{H}_{r,2}^{(t)}, \\
w_f^{(t)} &= (N_s p_1 \sigma_{e,1}^2 + \sigma_{n,1}^2) \text{tr}(\mathbf{R}_1^{(t)} \mathbf{R}_1^{(t)H}) + (N_s p_2 \sigma_{e,2}^2 + \sigma_{n,2}^2) \text{tr}(\mathbf{R}_2^{(t)} \mathbf{R}_2^{(t)H}).
\end{aligned} \tag{26}$$

For the Lagrangian in (25), we have three conditions of optimality as $\frac{\partial \mathcal{L}(\bar{\mathbf{F}}^{(t)}, \alpha^{(t)}, \lambda^{(t)})}{\partial \bar{\mathbf{F}}^{(t)}} = \mathbf{0}_{N_r}$, $\frac{\partial \mathcal{L}(\bar{\mathbf{F}}^{(t)}, \alpha^{(t)}, \lambda^{(t)})}{\partial \alpha^{(t)}} = 0$, and $\frac{\partial \mathcal{L}(\bar{\mathbf{F}}^{(t)}, \alpha^{(t)}, \lambda^{(t)})}{\partial \lambda^{(t)}} = 0$. Using the linear and nonlinear properties of the complex matrix derivative [21] and the cyclic permutation and linearity of the trace function, these conditions are reduced to

$$\mathbf{W}_{f,1}^{(t)} \bar{\mathbf{F}}^{(t)} \mathbf{G}_1^{(t)} + \mathbf{W}_{f,2}^{(t)} \bar{\mathbf{F}}^{(t)} \mathbf{G}_2^{(t)} + \lambda^{(t)} \alpha^{(t)^2} \bar{\mathbf{F}}^{(t)} \mathbf{G}_r^{(t)} = \mathbf{W}_{f,0}^{(t)} \quad (27)$$

$$\text{tr}(\bar{\mathbf{F}}^{(t)} \mathbf{G}_r^{(t)} \bar{\mathbf{F}}^{(t)H}) = \lambda^{(t)-1} \alpha^{(t)-4} w_f^{(t)} \quad (28)$$

$$\text{tr}(\bar{\mathbf{F}}^{(t)} \mathbf{G}_r^{(t)} \bar{\mathbf{F}}^{(t)H}) = \alpha^{(t)-2} N_r p_r \quad (29)$$

From (28) and (29), we have

$$\lambda^{(t)} \alpha^{(t)^2} = (N_r p_r)^{-1} w_f^{(t)} \quad (30)$$

and substituting (30) into (27), we obtain

$$\mathbf{W}_{f,1}^{(t)} \bar{\mathbf{F}}^{(t)} \mathbf{G}_1^{(t)} + \mathbf{W}_{f,2}^{(t)} \bar{\mathbf{F}}^{(t)} \mathbf{G}_2^{(t)} + (N_r p_r)^{-1} w_f^{(t)} \bar{\mathbf{F}}^{(t)} \mathbf{G}_r^{(t)} = \mathbf{W}_{f,0}^{(t)}. \quad (31)$$

To solve (31), we use the following lemma from [13].

Lemma 2 ([13]): Let $\mathbf{A}_{1,i} \in \mathbb{C}^{N_1 \times N_1}$, $\mathbf{A}_{2,i} \in \mathbb{C}^{N_2 \times N_2}$, and $\mathbf{A}_3 \in \mathbb{C}^{N_1 \times N_2}$ for $1 \leq i \leq k$. A matrix $\mathbf{X} \in \mathbb{C}^{N_1 \times N_2}$ is a solution of the general linear matrix equation.

$$\mathbf{A}_{1,1} \mathbf{X} \mathbf{A}_{2,1} + \mathbf{A}_{1,2} \mathbf{X} \mathbf{A}_{2,2} + \cdots + \mathbf{A}_{1,k} \mathbf{X} \mathbf{A}_{2,k} = \mathbf{A}_3 \quad (32)$$

if and only if $\mathbf{x} = \text{vec}(\mathbf{X})$ is a solution of the equation $\mathbf{K} \mathbf{x} = \mathbf{a}_3$ with $\mathbf{K} = \sum_{i=1}^k (\mathbf{A}_{2,i}^T \otimes \mathbf{A}_{1,i})$ and $\mathbf{a}_3 = \text{vec}(\mathbf{A}_3)$.

Using *Lemma 2*, the optimal $\bar{\mathbf{f}}^{(t)\dagger} \left(= \text{vec}(\bar{\mathbf{F}}^{(t)\dagger}) \right)$ is expressed as

$$\bar{\mathbf{f}}^{(t)\dagger} = \mathbf{W}_f^{(t)-1} \mathbf{w}_{f,0}^{(t)} \quad (33)$$

where $\mathbf{W}_f^{(t)} = (\mathbf{G}_1^{(t)T} \otimes \mathbf{W}_{f,1}^{(t)}) + (\mathbf{G}_2^{(t)T} \otimes \mathbf{W}_{f,2}^{(t)}) + \{\mathbf{G}_r^{(t)T} \otimes (N_r p_r)^{-1} w_f^{(t)} \mathbf{I}_{N_r}\}$ and $\mathbf{w}_{f,0}^{(t)} = \text{vec}(\mathbf{W}_{f,0}^{(t)})$.

Note that the optimal $\bar{\mathbf{F}}^{(t)}$ in time slot t involves the past channels from time slot $t - m$ to time

slot t since it is a function of $\mathbf{G}_1^{(t)}$, $\mathbf{G}_2^{(t)}$, and $\mathbf{G}_r^{(t)}$ in (26). Once the optimal $\bar{\mathbf{F}}^{(t)}$ is obtained, the optimal $\alpha^{(t)\dagger}$ can be obtained from (29) such that

$$\alpha^{(t)\dagger} = \{N_r p_r \text{tr}(\bar{\mathbf{F}}^{(t)} \mathbf{G}_r^{(t)} \bar{\mathbf{F}}^{(t)H})^{-1}\}^{1/2} \quad (34)$$

Substituting (34) into (30), the optimal $\lambda^{(t)\dagger}$ is given by

$$\lambda^{(t)\dagger} = (N_r p_r)^{-2} w_f^{(t)} \text{tr}(\bar{\mathbf{F}}^{(t)} \mathbf{G}_r^{(t)} \bar{\mathbf{F}}^{(t)H}) \quad (35)$$

Then, the optimal relay beamforming matrix in time slot t is obtained as $\mathbf{F}^{(t)\dagger} = \alpha^{(t)\dagger} \text{mat}(\bar{\mathbf{F}}^{(t)\dagger})$.

In the next time slot, the optimal relay beamforming matrix is derived in the same way.

Remark: The proposed relay beamforming design method can be applied to one-way FD AF relay systems. Unlike the two-way FD AF relay systems, using matrix inverse operation, the relay beamforming matrix solution can be obtained without *Lemma 2*.

B. Receive beamforming design at sources

The problem of receive beamforming matrices design in each time slot is formulated as

$$\mathbb{P}_r^{(t)} : \min_{\mathbf{R}_1^{(t)}, \mathbf{R}_2^{(t)}} J(\mathbf{R}_1^{(t)}, \mathbf{R}_2^{(t)}) \quad (36)$$

where the sum of MSE $J(\mathbf{R}_1^{(t)}, \mathbf{R}_2^{(t)})$ is defined as

$$\begin{aligned} J(\mathbf{R}_1^{(t)}, \mathbf{R}_2^{(t)}) = & N_s(p_1 + p_2) \\ & - \alpha^{(t)-1} \{p_2 \text{tr}(\mathbf{W}_{r,1}^{(t)H} \mathbf{R}_1^{(t)}) + p_2 \text{tr}(\mathbf{W}_{r,1}^{(t)} \mathbf{R}_1^{(t)H}) + p_1 \text{tr}(\mathbf{W}_{r,2}^{(t)H} \mathbf{R}_2^{(t)}) + p_1 \text{tr}(\mathbf{W}_{r,2}^{(t)} \mathbf{R}_2^{(t)H})\} \\ & + \alpha^{(t)-2} \{\text{tr}(\mathbf{W}_{r,3}^{(t)} \mathbf{R}_1^{(t)} \mathbf{R}_1^{(t)H}) + \text{tr}(\mathbf{W}_{r,4}^{(t)} \mathbf{R}_2^{(t)} \mathbf{R}_2^{(t)H})\} \end{aligned} \quad (37)$$

 Algorithm 1. The proposed iterative algorithm in time slot t

- 1: Initialize: $\mathbf{F}^{(t)} = \mathbf{I}_{N_r}$ and $\mathbf{R}_i^{(t)} = \mathbf{I}_{N_s}$ for $i \in \{1, 2\}$
 - 2: Repeat
 - 3: Update $\bar{\mathbf{F}}^{(t)}$ by using $\mathbf{R}_i^{(t)}$ for $i \in \{1, 2\}$
 - 4: Update $\alpha^{(t)}$ by using $\bar{\mathbf{F}}^{(t)}$
 - 5: Update $\mathbf{F}^{(t)}$ by using $\bar{\mathbf{F}}^{(t)}$ and $\alpha^{(t)}$
 - 6: Update $\mathbf{R}_i^{(t)}$ for $i \in \{1, 2\}$ by using $\mathbf{F}^{(t)}$
 - 7: Update $J(\mathbf{F}^{(t)}, \alpha^{(t)}, \mathbf{R}_1^{(t)}, \mathbf{R}_2^{(t)})$
 - 8: Until convergence.
-

where

$$\begin{aligned}
 \mathbf{W}_{r,1}^{(t)} &= \mathbf{H}_{r,1}^{(t)} \mathbf{F}^{(t)} \mathbf{H}_{2,r}^{(t-1)H}, \\
 \mathbf{W}_{r,2}^{(t)} &= \mathbf{H}_{r,2}^{(t)} \mathbf{F}^{(t)} \mathbf{H}_{1,r}^{(t-1)H}, \\
 \mathbf{W}_{r,3}^{(t)} &= \mathbf{H}_{r,1}^{(t)} \mathbf{F}^{(t)} \mathbf{G}_1^{(t)} \mathbf{F}^{(t)H} \mathbf{H}_{r,1}^{(t)H} + (N_s p_1 \sigma_{e,1}^2 + \sigma_{n,1}^2) \mathbf{I}_{N_s}, \\
 \mathbf{W}_{r,4}^{(t)} &= \mathbf{H}_{r,2}^{(t)} \mathbf{F}^{(t)} \mathbf{G}_2^{(t)} \mathbf{F}^{(t)H} \mathbf{H}_{r,2}^{(t)H} + (N_s p_2 \sigma_{e,2}^2 + \sigma_{n,2}^2) \mathbf{I}_{N_s}.
 \end{aligned} \tag{38}$$

By solving $\frac{\partial J(\mathbf{R}_1^{(t)}, \mathbf{R}_2^{(t)})}{\partial \mathbf{R}_i^{(t)}} = \mathbf{0}_{N_s}$ for $i \in \{1, 2\}$, the receive beamforming matrices at the sources are given by

$$\mathbf{R}_l^{(t)} = \alpha^{(t)} p_l \{ \mathbf{H}_{r,l}^{(t)} \mathbf{F}^{(t)} \mathbf{G}_l^{(t)} \mathbf{F}^{(t)H} \mathbf{H}_{r,l}^{(t)H} + (N_s p_l \sigma_{e,l}^2 + \sigma_{n,l}^2) \mathbf{I}_{N_s} \}^{-1} \mathbf{H}_{r,l}^{(t)} \mathbf{F}^{(t)} \mathbf{H}_{\bar{l},r}^{(t-1)} \tag{39}$$

where \bar{l} means the index of the other source ($\bar{l} = 2$ if $l = 1$ and $\bar{l} = 1$ if $l = 2$).

C. An iterative algorithm for joint beamforming design

An iterative algorithm is proposed for the joint relay and receive beamforming design. The relay beamforming matrix and receive beamforming matrices are optimized alternately until convergence in each time slot. The proposed iterative algorithm is shown in Algorithm 1.

The achievable sum rate in time slot t is given by

$$R^{(t)} = \mathbb{E}[\log_2 \det\{\mathbf{I}_{N_s} + p_1 \mathbf{R}_2^{(t)H} \mathbf{H}_{r,2}^{(t)} \bar{\mathbf{F}}^{(t)} \mathbf{H}_{1,r}^{(t-1)} \mathbf{H}_{1,r}^{(t-1)H} \bar{\mathbf{F}}^{(t)H} \mathbf{H}_{r,2}^{(t)} \mathbf{R}_2^{(t)} (\mathbf{R}_2^{(t)H} \mathbf{A}_2^{(t)} \mathbf{R}_2^{(t)})^{-1}\}] \\ + \mathbb{E}[\log_2 \det\{\mathbf{I}_{N_s} + p_2 \mathbf{R}_1^{(t)H} \mathbf{H}_{r,1}^{(t)} \bar{\mathbf{F}}^{(t)} \mathbf{H}_{2,r}^{(t-1)} \mathbf{H}_{2,r}^{(t-1)H} \bar{\mathbf{F}}^{(t)H} \mathbf{H}_{r,1}^{(t)} \mathbf{R}_1^{(t)} (\mathbf{R}_1^{(t)H} \mathbf{A}_1^{(t)} \mathbf{R}_1^{(t)})^{-1}\}] \quad (40)$$

where $\mathbf{A}_l^{(1)} = \sigma_{n,r}^2 \mathbf{H}_{r,l}^{(1)} \bar{\mathbf{F}}^{(1)} \bar{\mathbf{F}}^{(1)H} \mathbf{H}_{r,l}^{(1)H} + \alpha^{(1)-2} p_l \Delta_{l,l}^{(1)} \Delta_{l,l}^{(1)H} + \alpha^{(1)-2} \sigma_{n,l}^2 \mathbf{I}_{N_s}$ and

$$\mathbf{A}_l^{(t)} = \sigma_{n,r}^2 \mathbf{H}_{r,l}^{(t)} \bar{\mathbf{F}}^{(t)} \bar{\mathbf{F}}^{(t)H} \mathbf{H}_{r,l}^{(t)H} + \sum_{i=2}^t \left\{ \mathbf{H}_{r,l}^{(t)} \bar{\mathbf{F}}^{(t)} \prod_{j=1}^{i-1} (\Delta_{r,r}^{(t-j)} \mathbf{F}^{(t-j)}) \right. \\ \left. (p_l \mathbf{H}_{l,r}^{(t-i)} \mathbf{H}_{l,r}^{(t-i)H} + p_{\bar{l}} \mathbf{H}_{\bar{l},r}^{(t-i)} \mathbf{H}_{\bar{l},r}^{(t-i)H} + \sigma_{n,r}^2 \mathbf{I}_{N_r}) \prod_{j=t+1-i}^{t-1} (\mathbf{F}^{(j)H} \Delta_{r,r}^{(j)H}) \bar{\mathbf{F}}^{(t)H} \mathbf{H}_{r,l}^{(t)H} \right\} \quad (41) \\ + \alpha^{(t)-2} p_l \Delta_{l,l}^{(t)} \Delta_{l,l}^{(t)H} + \alpha^{(t)-2} \sigma_{n,l}^2 \mathbf{I}_{N_s}$$

and \bar{l} means the index of the other source ($\bar{l} = 2$ if $l = 1$ and $\bar{l} = 1$ if $l = 2$).

The sum of MSE $J_m^{(t)} = J(\mathbf{F}^{(t)}, \alpha^{(t)}, \mathbf{R}_1^{(t)}, \mathbf{R}_2^{(t)})$ depends on m and t . When $m = \infty$, $J_m^{(t)}$ increases or does not change as t increases. When m is a finite number, $J_m^{(t)}$ increases or does not change as t increases only if $t \leq m + 1$, but $J_m^{(t)}$ decreases as t increases if $t \geq m + 2$ because the relay substitutes the oldest beamforming matrix with a new one. In other words, the MSE of system can be unstable after time slot $m + 2$. Based on this observation, we propose an algorithm to determine m in Algorithm 2, where the determined m is denoted by \hat{m} . For example, if $J_{m=2}^{(3)} \leq J_{m=2}^{(4)}$, we set \hat{m} to be 2 for stability.

IV. NUMERICAL RESULTS

In this section, we numerically evaluate the sum of MSE and the achievable sum rate for the proposed scheme, conventional scheme, and the HD two-way relaying scheme. The proposed scheme updates the relay beamforming matrix $\mathbf{F}^{(t)}$ and receive beamforming matrix $\mathbf{R}_i^{(t)}$ for $i \in \{1, 2\}$ every time slot, which is based on the channels in the m latest time slots. Unlike the proposed scheme, the conventional scheme updates the relay beamforming matrix based only on

 Algorithm 2. The proposed algorithm for obtaining \hat{m}

```

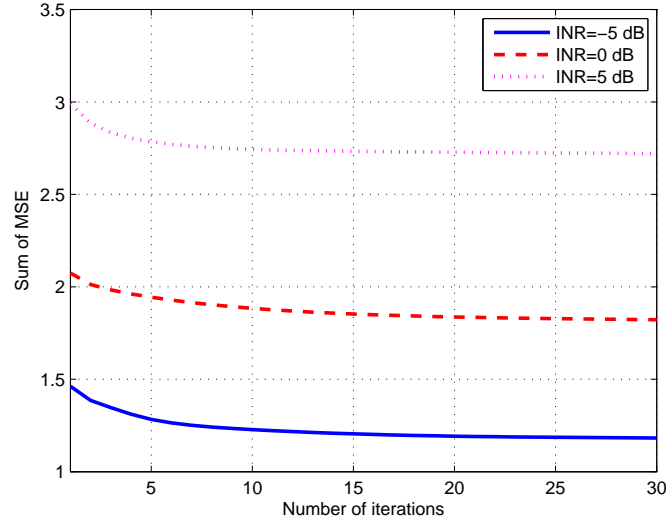
1:   Initialize:  $i = 0$ 
2:   Repeat
3:      $i \leftarrow i + 1$ 
4:     If  $J_{m=i}^{(i+1)} \leq J_{m=i}^{(i+2)}$ 
5:        $\hat{m} = i$ 
6:     End
7:   Until  $\hat{m}$  for stability is determined.

```

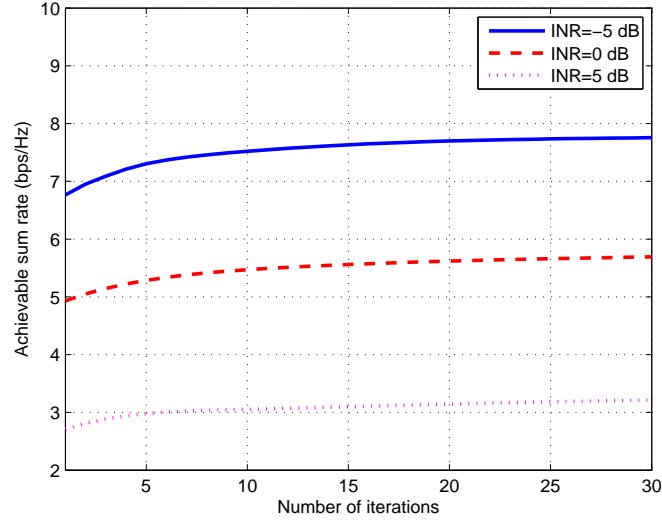
the channels in the current time slot, i.e. $\mathbf{G}_c^{(t)} = \mathbf{0}_{N_r}$. The HD two-way relaying scheme is used as a referential scheme. The sum of MSE and achievable sum rate are obtained by averaging over 100 channel realizations. The number of iteration for relay and receive beamforming design is 30.

The channel matrices $\mathbf{H}_{i,j}^{(t)}$ for $(i, j) \in \{(1, r), (r, 1), (2, r), (r, 2)\}$ are set to follow Rayleigh fading, i.e., the elements of each channel matrix are independent complex Gaussian random variables with zero mean and unit variance. The channel estimation error matrices $\Delta_{i,i}^{(t)}$ for $i \in \{1, 2, r\}$ are set to follow Rayleigh fading, i.e., the elements of each channel matrix are independent complex Gaussian random variables with zero mean and variance $\sigma_{e,i}^2$. For simplicity, it is assumed that all the nodes transmit with power p , i.e., the power is set to be $p_1 = p_2 = p_r = p = 1$. The noise variance and the loopback channel estimation error variance at each node are set to be as $\sigma_{n,1}^2 = \sigma_{n,2}^2 = \sigma_{n,r}^2 = \sigma_n^2$ and $\sigma_{e,1}^2 = \sigma_{e,2}^2 = \sigma_{e,r}^2 = \sigma_e^2$. SNR and INR are defined as p/σ_n^2 and σ_e^2/σ_n^2 , respectively.

Fig. 2 shows the sum of MSE and the achievable sum rate versus number of iterations for different INR values in the 10th time slot when SNR= 5dB, $N_s = 2$, $N_r = 5$, and $m = \infty$, respectively. The proposed scheme addresses two sub-problems for relay beamforming design



(a) Sum of MSE versus number of iterations for different INR values.

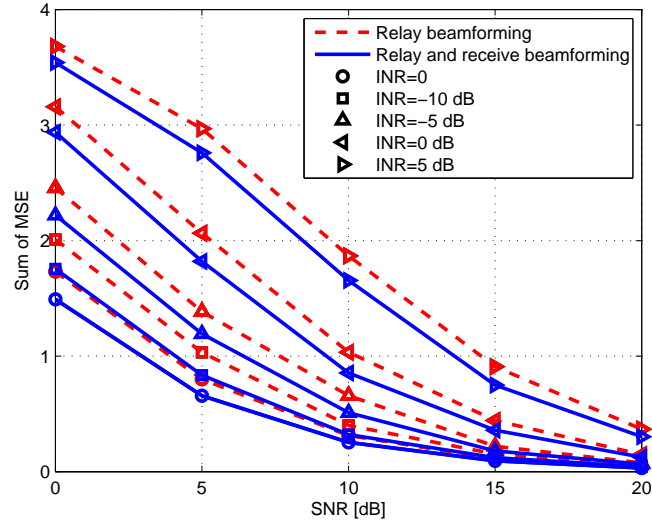


(b) Achievable sum rate versus number of iterations for different INR values.

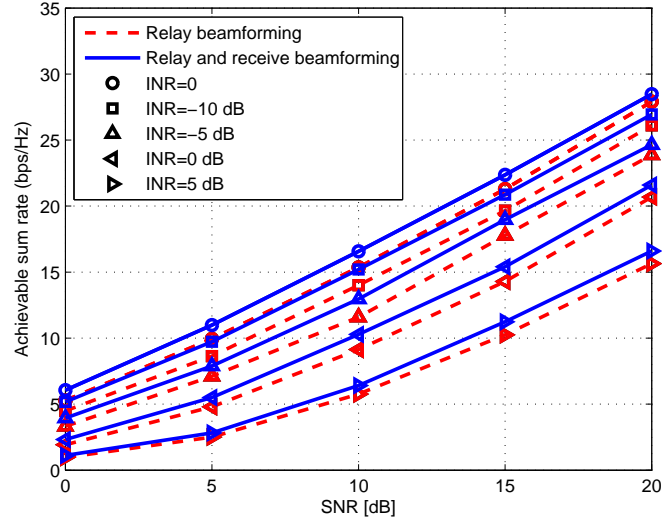
Fig. 2. Performance of the proposed scheme versus number of iterations for different INR

and receive beamforming design at sources. Since the solution for each sub-problem is optimal, the sum of MSE decreases with each iteration.

Fig. 3 shows the sum of MSE and achievable sum rate versus SNR for different INR values in



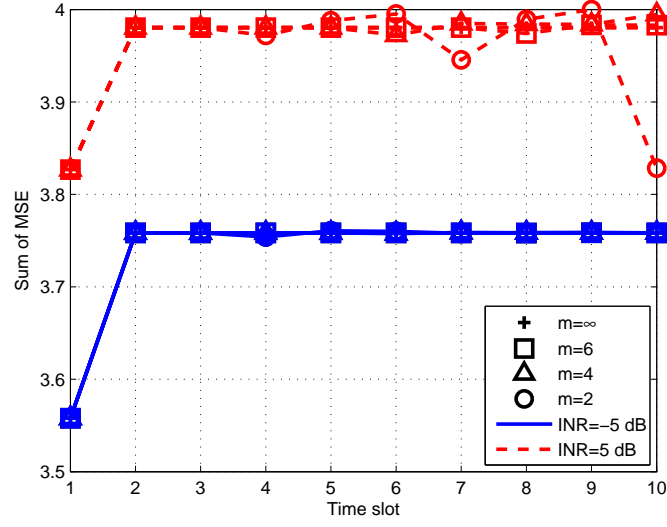
(a) Sum of MSE versus SNR for different INR values.



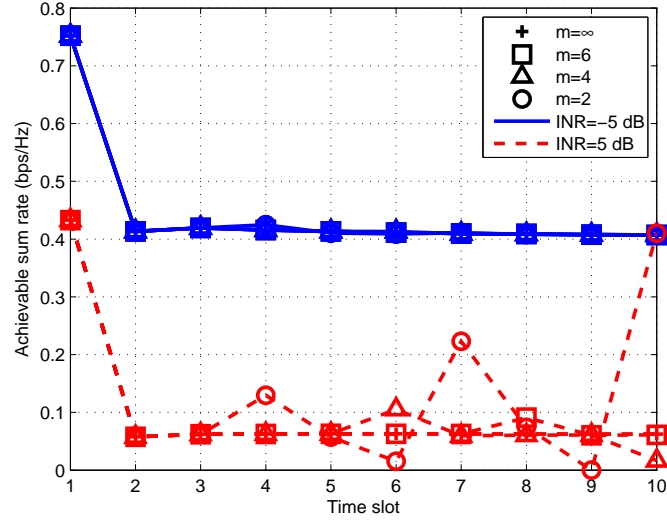
(b) Achievable sum rate versus SNR for different INR values.

Fig. 3. Performance comparisons between the proposed relay beamforming and the proposed joint beamforming schemes versus SNR

the 10th time slot when $N_s = 2$, $N_r = 5$, and $m = \infty$, respectively. As shown, the proposed joint beamforming design achieves better performance than the proposed design of relay beamforming only. As SNR increases and INR decreases, the sum of MSE decreases and the achievable sum



(a) Sum of MSE versus time slot index for different values of m and INR



(b) Achievable sum rate versus time slot index for different values of m and

INR

Fig. 4. Performance of the proposed joint beamforming design scheme versus time slot index for different values of m and INR

rate increases for both the proposed schemes.

Fig. 4 shows the sum of MSE and the achievable sum rate versus the time slot index for

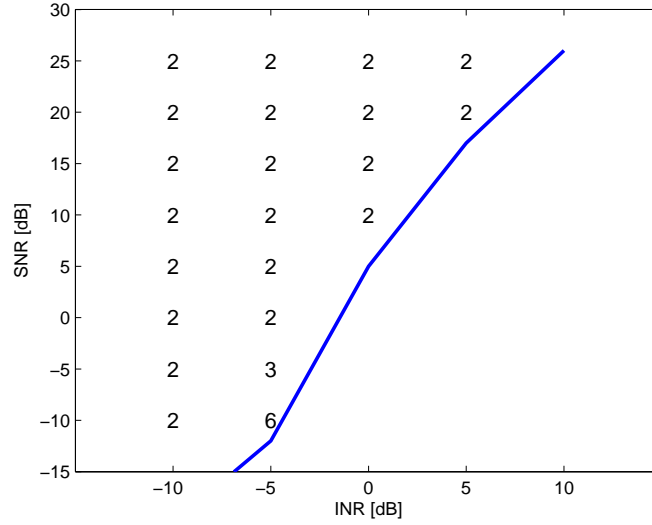


Fig. 5. The criterion for duplex mode selection and \hat{m} for stability in the FD mode.

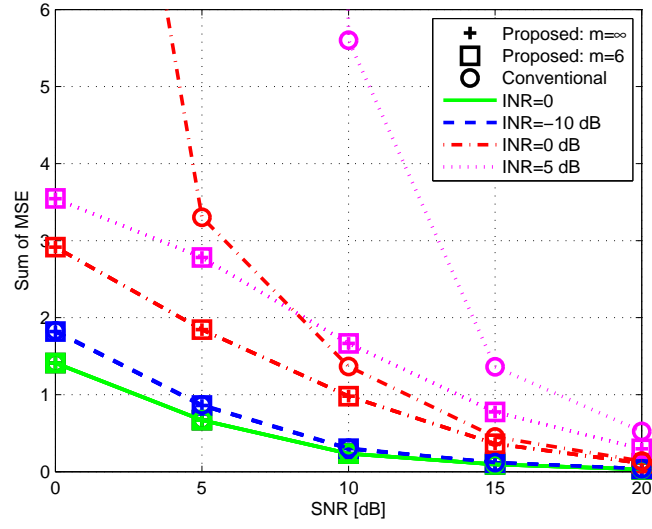
different values of m and INR when $\text{SNR} = -10$ dB, $N_s = 2$, and $N_r = 5$. The proposed scheme with $m = \infty$ and $m = \hat{m}$ are the methods that the relay and receive beamforming matrices are designed with all the beamforming and channel matrices from the first time slot and with only those in the \hat{m} latest time slots, respectively. If \hat{m} is too small, the performance is unstable after time slot $\hat{m} + 2$ because it cannot handle the propagated effect of the residual loopback SI. For instance, when $m = 2, 4$ under $\text{INR} = -5$ dB and $m = 2, 4, 6$ under $\text{INR} = 5$ dB, both of the sum MSE and the sum rate oscillate and the oscillations become larger as either INR increases or m decreases. On the other hand, if $m \geq 6$ under $\text{INR} = -5$ dB, the performance gap from the case of $m = \infty$ is marginal.

Fig. 5 shows the criteria for the duplex mode selection and the selection of \hat{m} for stability in the FD mode when $N_s = 2$ and $N_r = 5$. In the region above the solid line, the achievable sum rate of the proposed scheme in FD mode is greater than that of HD mode, so FD is preferred to HD. However, below the solid line, HD is better than FD. This figure also reveals that as

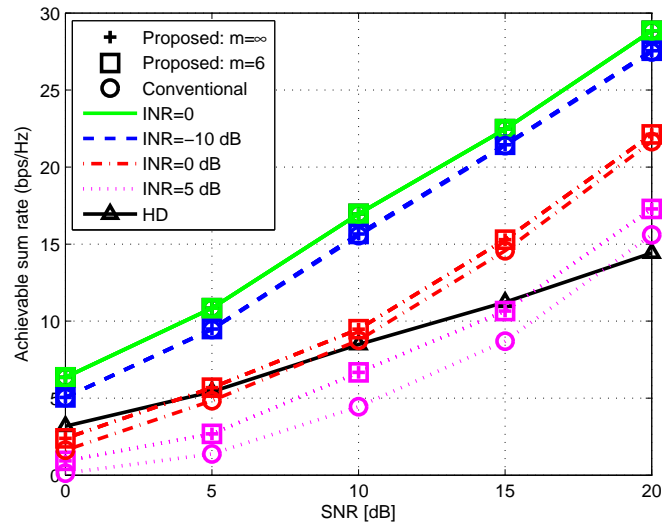
INR increases, the SNR required for selection of FD grows. For each pair of SNR and INR, the values of \hat{m} for stability are presented, too. For the region where HD is preferred, \hat{m} is not specified because the residual loopback SI in the HD mode does not exist.

Fig. 6(a) shows the sum of MSE versus SNR for different INR values in the 10th time slot when $N_s = 2$ and $N_r = 5$. It is shown that the proposed scheme of joint beamforming design provides a lower MSE than the conventional scheme. As SNR increases and INR decreases, the MSE decreases for both the proposed and conventional schemes. The performance difference between the two schemes decreases as SNR increases and INR decreases. This is because the proposed scheme updates the relay and receive beamforming matrices based on the covariance matrix of loopback channel estimation error. This figure also verifies that $m = 6$ provides almost the same MSE performance as $m = \infty$. Fig. 6(b) compares the achievable sum rates of the proposed scheme of joint beamforming design and the conventional scheme at the 10th time slot when $N_s = 2$ and $N_r = 5$. In either low SNR or high INR regime, FD is not beneficial relative to HD because the cost of handling the residual loopback SI is higher than the HD loss in those regimes, which strongly suggests adaptive selection between HD and FD according to SNR and INR; if INR is less than 0 dB when $\text{SNR} \geq 5$ dB or less than 5 dB when $\text{SNR} \geq 16$ dB, FD with the proposed beamforming scheme is preferred to HD. If we focus on the scenarios where FD is preferred, $m = 6$ offers almost the same achievable sum rate as $m = \infty$, as in Fig. 6(a).

Fig. 7 shows the achievable sum rate gain versus SNR for different INR values at the 10th time slot when $N_s = 2$ and $N_r = 5$. The achievable sum rate gain is defined as the ratio between the achievable sum rate of the proposed joint design scheme and that of the HD scheme. For example, the achievable sum rate gain is 2 and 1.9 when $\text{INR} = 0$ and INR is -10 dB, respectively. The



(a) Sum of MSE versus SNR for different INR values.



(b) Achievable sum rate versus SNR for different INR values.

Fig. 6. Performance comparison between the proposed scheme of joint beamforming design and the conventional scheme versus SNR

gain increases as SNR decreases or INR increases.

Fig. 8 shows the sum of MSE and achievable sum rate versus the time slot index for different INR (i.e., $\text{INR} = -5$ dB, 0 dB, and 5 dB) when $\text{SNR} = 5$ dB, $N_s = 2$, and $N_r = 5$. The

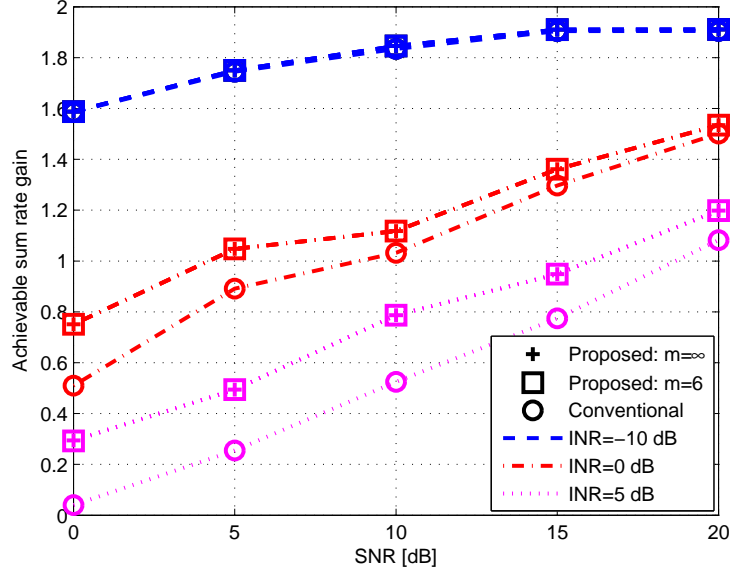
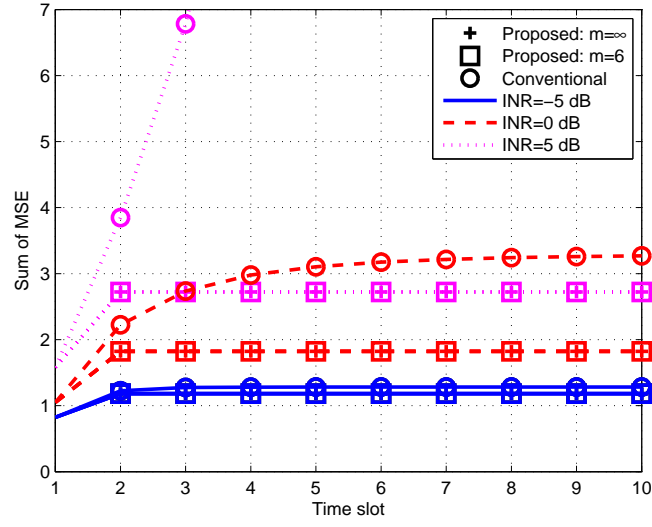


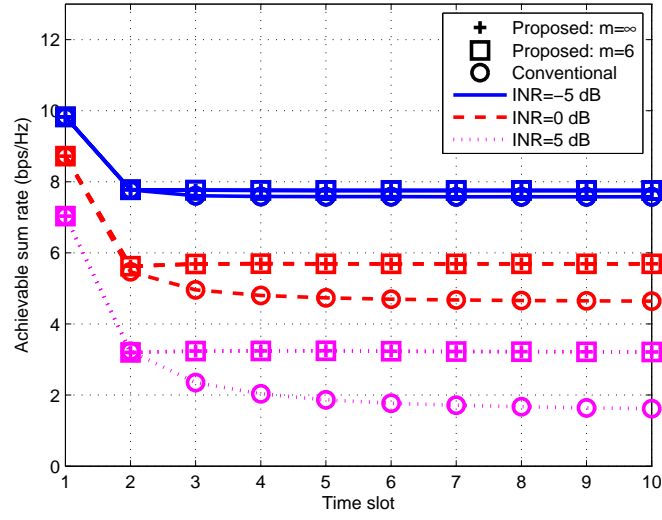
Fig. 7. Achievable sum rate gain versus SNR for different INR values.

performance gap between the proposed joint design scheme and the conventional scheme is shown to be larger as the time slot index increase, for both MSE and sum rate. The error propagation effect of residual loopback SI over time slots is not properly addressed in the conventional scheme and accordingly the loss due to the error propagation increases with time slot index. Interestingly, the performance gap between the first and the second time slots is greater than that between the second and the third time slots. This is because the residual SI at relay starts to affect from the second time slot whereas the residual SI at source affects from the first time slot. As in the previous figures, it is verified again that $m = 6$ offers almost the same performance as $m = \infty$.

Fig. 9 presents the sum of MSE and achievable sum rate versus SNR for different values of N_r at the 10th time slot when $N_s = 2$ and INR= 0 dB. As the N_r increases, the MSE decreases and achievable sum rate increases for both the proposed joint design scheme and the conventional



(a) Sum of MSE versus time slot index



(b) Achievable sum rate versus time slot index

Fig. 8. Performance comparison between the proposed and conventional schemes versus time slot index

scheme. The performance gap between the proposed and conventional schemes increases with N_r because the multiple antennas at relay are more efficiently used in the proposed scheme.

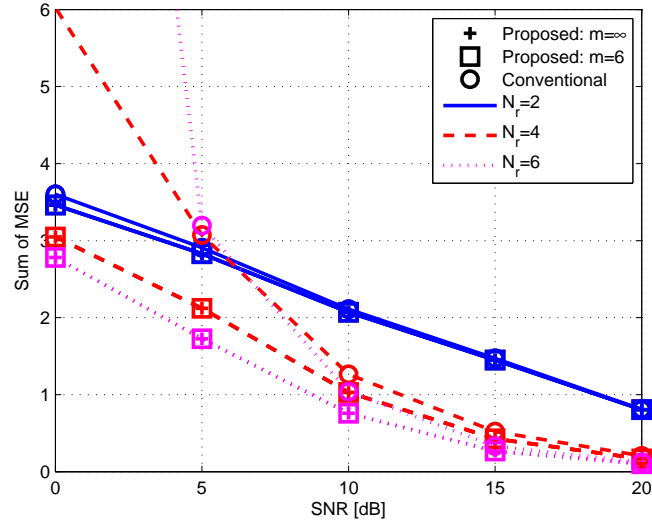
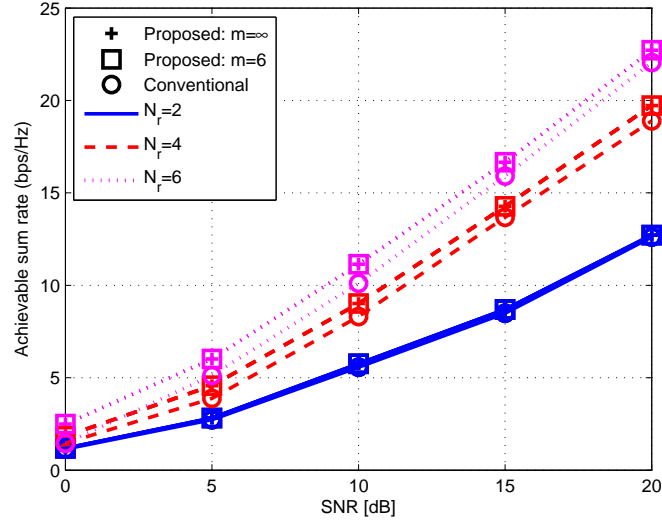
(a) Sum of MSE versus SNR for different values of N_r (b) Achievable sum rate versus SNR for different values of N_r

Fig. 9. Performances comparison between the proposed and conventional schemes versus SNR

V. CONCLUSION

This paper investigated joint design of relay beamforming and receive beamforming at sources in FD two-way AF relay systems under a relay transmit power constraint. We analyzed the cou-

pled effect of beamforming matrix design across time slots and derived the optimal beamforming matrices in closed form at each time slot in terms of MMSE, considering the coupled effect. To reduce the burden on the beamforming design incorporating all the time slots from the first time slot to the current time slot, we also proposed the beamforming design method based only on the m latest time slots. It was shown that if m is not too small, the performance degradation due to a limited number of m was marginal. With numerical analysis, we revealed when full-duplex operation is beneficial compared to HD operation, and proposed adaptive selection between FD and HD according to INR and SNR.

REFERENCES

- [1] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062-3080, Dec. 2004.
- [2] B. Rankov and A. Wittneben, "Achievable rate regions for the two-way relay channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, July 2006.
- [3] B. Rankov and A. Wittneben, "Spectral efficient protocols for half-duplex fading relay channels," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 379-389, Feb. 2007.
- [4] S. J. Kim, P. Mitran, and V. Tarokh, "Performance bounds for bidirectional coded cooperation protocols," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 5235-5241, Nov. 2008.
- [5] S. J. Kim, N. Devroye, P. Mitran, and V. Tarokh, "Achievable rate regions and performance comparison of half duplex bi-directional relaying protocols," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6405-6418, Oct. 2011.
- [6] F. S. Tabataba, P. Sadeghi, C. Hucher, and M. R. Pakravan, "Impact of channel estimation errors and power allocation on analog network coding and routing in two-way relaying," *IEEE Trans. Veh. Technol.*, vol. 61, no. 7, pp. 3223-3239, Sep. 2012.
- [7] X. Cheng, B. Yu, X. Cheng, and L. Yang, "Two-way full-duplex amplify-and-forward relaying," *IEEE Military Commun. Conf. (MILCOM)*, San Diego, USA, Nov. 2013.
- [8] Y. Li, T. Wang, Z. Zhao, M. Peng, and W. Wang, "Relay mode selection and power allocation for hybrid one-way/two-way half-duplex/full-duplex relaying," *IEEE Commun. Lett.*, vol. 19, no. 7, pp. 1217-1220, July 2015.

- [9] D. Choi and J. H. Lee, "Outage probability of two-way full-duplex relaying with imperfect channel state information," *IEEE Commun. Lett.*, vol. 18, no. 6, pp. 933-936, June 2014.
- [10] X. Cheng, B. Yu, X. Cheng, and L. Yang, "Practical full duplex physical layer network coding," *IEEE Vehicular Technol. Conf. (VTC)*, Seoul, Korea, May 2014.
- [11] H. Cui, M. Ma, L. Song, and B. Jiao, "Relay selection for two-way full duplex relay networks with amplify-and-forward protocol," *IEEE Trans. Wireless Commun.*, vol. 13, no. 7, pp. 3768-3777, July 2014.
- [12] G. Zheng, "Joint beamforming optimization and power control for full-duplex MIMO two-way relay channel," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 555-566, Feb. 2015.
- [13] G. Li, Y. Wang, and P. Zhang, "Optimal linear MMSE beamforming for two way multi-antenna relay systems," *IEEE Commun. Lett.*, vol. 15, no. 5, pp. 533-535, May 2011.
- [14] R. Wang and M. Tao, "Joint source and relay precoding designs for MIMO two-way relaying based on MSE criterion," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1352-1365, Mar. 2012.
- [15] W.-C. Choi, S. Choi, A. Heo, and D.-J. Park, "Joint optimization of source and relay for MIMO two-way relay networks using MSE duality," *IEEE Commun. Lett.*, vol. 18, no. 7, pp. 1246-1249, July 2014.
- [16] Z. He, Z. Lang, Y. Rong, and S. Qu, "Joint transceiver optimization for two-way MIMO relay systems with MSE constraints," *IEEE Wireless Commun. Lett.* vol. 3, no. 6, pp. 613-616, Dec. 2014.
- [17] H. Park, C. Song, H. Lee, and I. Lee, "MMSE-based filter designs for cognitive multiuser two-way relay networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1638-1643, Apr. 2015.
- [18] K. X. Nguyen, Y. Rong, and S. Nordholm, "Simplified MMSE precoding design in interference two-way MIMO relay systems," *IEEE Signal Process. Lett.*, vol. 23, no. 2, pp. 262-266, Feb. 2016.
- [19] J. Joung and A. H. Sayed, "Multiuser two-way amplify-and-forward relay processing and power control methods for beamforming systems," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1833-1846, Mar. 2010.
- [20] P. Ubaidulla and A. Chockalingam, "Relay precoder optimization in MIMO-relay networks with imperfect CSI," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5473-5484, Nov. 2011.
- [21] A. Hjørungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2740-2746, May 2007.